REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

AD-A248 481

Pub... gat... colle... Dav... age 1 hour per response, including the time for reviewing instructions, searching existing data sources, ...e collection of information. Send comments regarding this burden estimate or any other aspect of this ...Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson ...anagement and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. ... **E**

3. REPORT TYPE AND DATES COVERED
FINAL 15 Dec 88 - 14 Jun 91

4. TITLE AND SUBTITLE
"NUMERICAL METHODS FOR CLOSED-LOOP CONTROL" (U)

5. FUNDING NUMBERS
61102F
2304/A1

6. AUTHOR(S)
Professor Alan J. Laub

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of California
Electrical & Computer Engineering
Santa Barbara, CA 93106-9560

8. PERFORMING ORGANIZATION REPORT NUMBER
AFOSR-TR- 92 0185

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/NM
Bldg 410
Bolling AFB DC 20332-6448

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFOSR-89-0167

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for Public Release;
Distribution Unlimited

12b. DISTRIBUTION CODE
UL

13. ABSTRACT *(Maximum 200 words)*

The primary objective of this research has been to develop algorithms for large-scale computational problems arising in control, filtering, and system theory. Much work has concentrated on matrix Riccati equations which are absolutely fundamental, but a range of other topics in control design have also been attacked.

14. SUBJECT TERMS

15. NUMBER OF PAGES
26

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR

Numerical Methods for Closed-Loop Control

FINAL REPORT

Air Force Office of Scientific Research

Contract No. AFOSR-89-0167

Dec. 15, 1988 - Jun. 14, 1991

Principal Investigator:

Professor Alan J. Laub
Dept. of Electrical and Computer Engineering
University of California
Santa Barbara, CA 93106-9560

tel: (805) 893-3616; fax: (805) 893-3262
e-mail: laub@ece.ucsb.edu

1 Factual Data

This section contains a factual listing of publications, research lectures, and students supported in whole or in part under Contract No. AFOSR-89-0167 with total funding support of $128,035 during the period December 15, 1988 to June 14, 1991. In particular, our research advances have led to 15 open literature publications, including 12 in the leading IEEE and SIAM journals. They are listed below and the narrative to follow in Section 2 is keyed to this list. Significant progress made under the auspices of the contract as well as plans for the future will be documented in the narrative. Finally, copies of the abstracts of papers published or accepted for publication will follow in Section 3.

1.1 Publications Supported by this AFOSR Contract

[1] Gahinet, P., A.J. Laub, C. Kenney, and G. Hewer, "Sensitivity of the Stable Discrete-Time Lyapunov Equation," *IEEE Trans. Aut. Control*, AC-35(1990), 1209-1217.

[2] Pandey, P., C. Kenney, and A.J. Laub, "A Parallel Algorithm for the Matrix Sign Function," *Int. J. High Speed Computing*, 2(1990), 181-191.

[3] Gahinet, P., and A.J. Laub, "Computable Bounds for the Sensitivity of the Algebraic Riccati Equation," *SIAM J. Contr. Opt.*, 28(1990), 1461-1480.

[4] Pandey, P., C. Kenney, A.J. Laub, and A. Packard, "Algorithms for Computing the Optimal H_∞ Norm," *Proc. 29th IEEE Conf. on Decision and Control*, Honolulu, Hawaii; December 1990; pp. 2628-2629.

[5] Pandey, P., C. Kenney, and A.J. Laub, "Numerical Solution of Large-Scale Riccati Equations," *Proc. Third Rockwell Advanced Control Systems/Neural Network/Signal Processing Conf.*, Anaheim, California; January 1991; pp. 100-112.

[6] Kenney, C., and A.J. Laub, "Rational Iterative Methods for the Matrix Sign Function," *SIAM J. Matrix Anal. Appl.*, 12(1991), 273-291.

[7] Kenney, C., and A.J. Laub, "Polar Decomposition and Matrix Sign Function Condition Estimates," *SIAM J. Sci. Stat. Comp.*, 12(1991), 488-504.

[8] Roy, S., R.H. Hashemi, and A.J. Laub, "Square Root Parallel Kalman Filtering Using Reduced-Order Local Filters," *IEEE Trans. Aerosp. Electr. Sys.*, 27(1991), 276-289.

[9] Laub, A.J., "Invariant Subspace Methods for the Numerical Solution of Riccati Equations," in *The Riccati Equation*, S. Bittanti, A.J. Laub, and J.C. Willems (eds.), Springer-Verlag, Berlin, 1991, pp. 163-196.

[10] Pandey, P., C. Kenney, A. Packard, and A.J. Laub, "A Gradient Method for Computing the Optimal H_∞ Norm," *IEEE Trans. Aut. Contr.*, AC-36(1991), 887-890.

[11] Ghavimi, A., C. Kenney, and A.J. Laub, "Local Convergence Analysis of Conjugate Gradient Methods for Solving Algebraic Riccati Equations," to appear in *IEEE Trans. Aut. Contr.*, 1992.

[12] Gahinet, P., and A.J. Laub, "Algebraic Riccati Equations and the Distance to the Nearest Uncontrollable Pair," to appear in *SIAM J. Contr. Opt.*, 1992.

[13] Kenney, C., and A.J. Laub, "On Scaling Newton's Method for Polar Decomposition and the Matrix Sign Function," to appear in *SIAM J. Matrix Anal. Appl.*, 1992 (early version also appeared as "On Scaling Newton's Method for Polar Decomposition and the Matrix Sign Function," *Proc. 1990 American Control Conf.*, San Diego, California; May 1990; pp. 2560-2564).

[14] Gudmundsson, T., C. Kenney, and A.J. Laub, "Scaling of the Discrete-Time Algebraic Riccati Equation to Enhance Stability of the Schur Solution Method," to appear in *IEEE Trans. Aut. Contr.*, 1992.

[15] Williams, T., and A.J. Laub, "Orthogonal Canonical Forms for Second-Order Systems," to appear in *IEEE Trans. Aut. Contr.*, 1992 (early version also appeared as "Orthogonal Canonical Forms for Second-Order Systems," *Proc. 1989 American Control Conf.*, Pittsburgh, Pennsylvania; June 1989; pp. 1621-1622).

1.2 Major Invited Talks and Addresses Supported by this AFOSR Contract

1. *State Space Computing: Past, Present, and Future*: **Plenary Lecture** for the SIAM Conference, "Control in the Nineties: Achievements, Opportunities, and Challenges," San Francisco, CA, May 17-19, 1989; also given as a seminar at the Dept. of Mechanical Engineering, University of California, Irvine, Apr. 21, 1989.

2. *The Matrix Sign Function and Riccati Equations*: Systems Research Center, University of Maryland, College Park, MD, Apr. 27, 1989.

3. *Numerical Techniques for the Solution of Riccati Equations*: Invited **Plenary Tutorial Lecture** for the Workshop on "The Riccati Equation in Control, Systems, and Signals," Como, Italy, Jun. 26-28, 1989.

4. *Control Algorithms and Software Survey*: Invited **Plenary Lecture** for the 3rd Annual Conference on Aerospace Computational Control, Oxnard, California, Aug. 28-30, 1989.

5. *Riccati Equations and the Matrix Sign Function*: RIACS, NASA Ames, Moffett Field, CA, Oct. 17, 1989; Dept. of Electrical Engineering, Princeton University, Princeton, NJ, Oct. 25, 1989; Dept. of Electrical Engineering, University of Illinois, Urbana, IL, Nov. 3, 1989.

6. *Computational Problems in Control Theory*: Dept. of Electrical Engineering, University of Pennsylvania, Philadelphia, PA, Oct. 26, 1989.

7. *The Matrix Sign Function and Large-Scale Riccati Equations*: Berkeley Center for Systems and Control, Spring Seminar Series, University of California, Berkeley, CA, May 9, 1990; Invited **Plenary Lecture** for SIAM Annual National Meeting, Chicago, Illinois, July 18, 1990.

8. *IEEE Control Systems Society Distinguished Lecture Series — Numerical Linear Algebra Problems in Control Theory*: Ohio State University, Columbus, OH, October 22, 1990; Wright State University, Dayton, OH, October 23, 1990.

1.3 Graduate Students Supported by this AFOSR Contract

1.3.1 Ph.D. Dissertations Completed

1. **Pascal M. Gahinet**
   - *Perturbational and Topological Aspects of Sensitivity in Control Theory*
   - December 1989
   - Presently: Research Scientist, INRIA, Domaine de Voluceau, Le Chesnay (Paris), France
2. **Pradeep Pandey**
   - *Numerical Algorithms for Robust Control Problems*
   - December 1990
   - Presently: Research Scientist, Integrated Systems, Inc., Santa Clara, California

1.3.2 Other Graduate Students Supported and Expected Completion Data

1. Thorkell T. Gudmundsson (Ph.D., Aug. 1992)
2. Ali R. Ghavimi (Ph.D., Dec. 1992)
3. Philip Papadopoulos (Ph.D., Dec. 1992)

Citizenship: Of the 5 graduate students listed above, 3 are U.S. citizens. Gahinet (France) held an F-1 student visa as does Gudmundsson (Iceland). In addition to the above, four other Ph.D. students also work in the P.I.'s research group: Thomas A. Bryan (Aug. 1992), Mark A. Erickson (Jun. 1993), John J. Hench (Aug. 1992), and Stephen C. Stubberud (Aug. 1992). All are supported by other contracts or fellowships and all are U.S. citizens.

It should also be mentioned that the P.I. and his students benefit enormously by being members of the Center for Control Engineering and Computation at UCSB. This Center, under the co-directorship of Professor Petar Kokotović and the P.I., consists of nine permanent members from the Departments of Electrical and Computer Engineering, Chemical and Nuclear Engineering, and Mechanical and Environmental Engineering. The cross-departmental nature of the Center gives it great strength for its most important task which is to initiate and coordinate research projects rich in opportunities for cross-disciplinary investigations and applications to industrial, environmental, and defense systems. In contacts with industry, the Center benefits from an unusually rich experience of its members, covering a wide range of technologies.

2 Narrative

In this section we shall highlight research progress made under this AFOSR Contract and give some indication of current directions of research. Of course, a more detailed description of some of these research topics is contained in the original proposal.

The primary objective of this project has been the study of algorithms for large-scale computational problems arising in control, filtering, and system theory. Much of our work has concentrated on matrix Riccati equations which are absolutely fundamental to the field. Substantial progress has been made in other areas as well and we give the highlights of some of the more exciting contributions below with further narrative to follow.

(1): new parallel algorithms and successful implementations for several key computational problems in control and filtering

(2): significant breakthroughs in the area of error and condition estimation for the algebraic Riccati equation

(3): development of an entire family of iterative methods, together with a complete convergence analysis, for computing the matrix sign function; this family of methods is particularly effective for the solution of large-scale invariant subspace calculations because of its amenability to implementation on parallel and vector computers

(4): the first computationally reliable method for estimating the conditioning of the matrix sign function based on the Fréchet derivative; this work reveals fascinating parallels between the conditioning of the Riccati equation and the sign function of an associated Hamiltonian matrix

(5): a new scaling strategy for Newton's method for finding the sign of a matrix; this method eliminates certain problems of determinantal scaling while remaining nearly optimal in terms of speed

(6): development and analysis of conjugate gradient methods for solving Riccati equations and general classes of matrix equations; necessary and sufficient conditions for convergence have been derived in terms of the invertibility of the associated Fréchet derivative

(7): a quadratically convergent gradient method of determining optimal H_∞ norms; the method is much faster than current bisection methods and can be extended to more general perturbation problems

(8): enhanced understanding of certain matrix "nearness" problems

(9): enhanced understanding of what can and can not be done using reliable numerical procedures for matrix second-order models

Each of these results will be described in more detail below.

(1): We have developed several new algorithms for parallel computers and enjoyed successful implementations of some of them. For example, we have published research in [8] on a number of strategies and hierarchies for implementing Kalman filters in a decentralized or parallel way. A major contribution here is to describe and analyze various multisensor network scenarios whose signal processing tasks are amenable to multiprocessor implementation. A number of extant strategies are unified and extended and new algorithms are proposed which have the potential for approximately linear speed-up, are reasonably failure-resistant, and are optimized with respect to communication bandwidth and memory requirements at the various processors in the architecture. A special feature of the principally suggested architecture is the ability to accommodate parallel local filters of smaller state dimension than the global filter. A significant innovation in [8] relative to previous work is the description of specific implementation details for so-called square-root versions of filters in both covariance and information filter forms. Another parallel algorithm for matrix Riccati equations is discussed in (3) below.

(2): We have succeeded in deriving and extending computable error bounds for the solution of the algebraic Riccati equation (ARE). These bounds are based on the theory developed by Kenney, Laub, and Wette (*Sys. Contr. Lett.*, 12(1989), 241-150 for the Schur method,

and *Math. Contr. Sig. Sys.*, 3(1990), 211-224 for Newton refinement). These results and other algorithms based on invariant subspaces are reviewed in an extensive survey paper [9] which includes over 230 references. The Schur and Newton results are complementary in the sense that the Schur method error bound, which is based on an invariant subspace perturbation result of Stewart (*SIAM Review*, 15(1973), 727-764), is needed to guarantee that the computed solution is within the region of convergence of the Newton refinement method. A combination of using the Schur method and error bound as a "starter" for Newton refinement with residual error estimation, can be easily implemented numerically and the necessary changes to existing software are minimal. These results have been extended to discrete and singular Riccati problems. For example, the discrete sensitivity results in [1] extend those obtained for continuous systems by Kenney and Hewer (*SIAM J. Contr. Opt.*, 28(1990), 50-69). Similar extensions of the Kenney, Laub, and Wette results will appear in [14]. Both theoretical and computable bounds are determined and we note that the discrete-time case turned out to be somewhat nontrivial to handle. We have also developed a much deeper general understanding of the sensitivity of Riccati equations. Based on the Ph.D. dissertation of Pascal Gahinet, we have addressed in [3] the problem of determining computable bounds for the condition or sensitivity of ARE's. Specifically, when solving for the unique symmetric nonnegative definite solution of an ARE in finite precision arithmetic, it is crucial to know topological properties of such a solution when the parameter matrices of the equation are subject to perturbation.

(3): Since the Schur method may be impractical for the very large Riccati problems which can arise, for example, in distributed parameter control systems, we have also made enormous progress in extending the matrix sign function approach. This work has its roots in the important extension of the matrix sign function to generalized eigenvalue problems developed by Gardiner and Laub (*Int. J. Control*, 44(1986), 823-832). Those algorithms are based on applying Newton's method to a simple matrix equation for computing the sign of a certain matrix and then solving a certain linear system. This has led further to the search for more efficient methods of evaluating the matrix sign function, and to the development of a major paper [6] based on Padé approximation of a certain hypergeometric function. This key paper introduces a new family of algorithms, of which the classical Newton iteration is but a special case. The algorithms are especially amenable to implementation on both parallel and vector computers and a complete numerical analysis, including global convergence results, is developed in [6]. As part of Pradeep Pandey's Ph.D. dissertation, vectorized and parallel versions of these algorithms have been implemented on a Cray Y-MP supercomputer at NASA Ames. For reference we note that even our early results have shown that a 100th-order Riccati equation can be solved on this machine in 0.8 sec. Admittedly, not everyone has access to a Cray — at this moment. However, it is important to bear in mind that we will soon have Cray-type computing available in desk-top workstations in the next few years in much the same way we presently have workstations with as much or more computing power than the industry-standard VAX computer of only a few years ago. Descriptions of Cray implementations of efficient parallel partial-fraction versions of high-order formulas from our new family of algorithms have been published in [2] and [5]. In fact, in [5] we discuss the numerical solution of Riccati equations of order 556 (involving Hamiltonian matrices of order 1112) in joint work with Rockwell's Rocketdyne Division. The problem derives from a model associated with Space Station Freedom in which 278 modes are included.

(4): Aside from matters of just efficiency, the important numerical question of the sensitivity of matrix sign solutions has also been considered. In [7] a reliable condition estimation procedure

is presented which costs two extra sign function evaluations. This work is motivated by some of the fundamental research by Kenney and Laub (*SIAM J. Matrix Anal. Appl.*, 10(1989), 191-209) on estimating condition of general matrix-valued functions. Future research efforts will attempt to reduce the cost of this condition estimate, and extend these finite-dimensional Riccati results to infinite-dimensional operator problems arising from distributed parameter control systems.

(5): Also associated with the matrix sign problem is the use of scaling factors to accelerate convergence. In analyzing optimal scaling factors for the related problem of accelerating Newton's method for polar decompositon, we have discovered that the commonly used determinantal scaling for the sign function can behave non-optimally in some rather ordinary situations. A novel scaling strategy based on the spectral radius is also flawed but the analysis in [13] shows that the strengths of these two procedures can be combined in such a way as to produce a simple and efficiently realizable scaling method which is almost always optimal. This research has, of course, significant practical value for these important matrix calculations.

(6): We have also made great strides in developing and analyzing conjugate gradient (cg) methods for solving Riccati equations and general classes of matrix equations. In particular, we have been able to show that the cg method converges in a neighborhood of a solution if and only if the Fréchet derivative of the matrix problem at the solution is invertible. This means, for example, that the cg method is convergent near the positive extremal solution of an ARE because the stability of the closed-loop system matrix ensures that the Fréchet derivative (which in this case is just the usual Lyapunov operator) is *nonsingular* at that solution. Our cg algorithms can be applied to both symmetric and nonsymmetric Riccati equations, including those in various "nonstandard" formats (e.g., certain additional terms). The methods can also be extended to a wide class of general nonlinear matrix-valued equations. This is a very promising approach for very large-scale problems and our first results will be published in [11].

(7): In a related development, a quadratically convergent gradient method for finding optimal H_∞ norms has been derived. Empirical evidence shows that this approach is much faster than current techniques such as bisection methods, and can easily be extended to more general H_∞ problems. These results have been published in [10]. Other new technical characterizations of Riccati solutions arising in the H_∞ problem appear in [4].

(8): A key paper [12] will soon be published in the area of matrix "nearness" problems. In this work, a thorough mathematical treatment is given of the key problem of determining the nearness to uncontrollability of a given controllable state-space model. The key tool used in the analysis is a connection between nearness to unstabilizability and the behavior of the unique symmetric positive definite stabilizing solution of an associated algebraic Riccati equation.

(9): Finally, an important result relating to the matrix triples commonly found in so-called matrix second-order models will appear in [15]. The basic idea is to establish which canonical forms are obtainable under orthogonal equivalence for the standard matrix triple consisting of a mass matrix, a stiffness matrix, and a damping matrix. Equivalence under orthogonal transformations is, of course, crucial for numerical reliability. It is established that an arbitrary damping model can not be used but that orthogonal reduction of the commonly used modal damping model can be so reduced.

Thus we continue to be excited about the progress that has been made and are extremely enthusiastic about the prospects and opportunities for further research.

3 Abstracts of Papers Published or Accepted for Publication

See attached.

Sensitivity of the Stable Discrete-Time Lyapunov Equation

PASCAL M. GAHINET, ALAN J. LAUB, FELLOW, IEEE, CHARLES S. KENNEY, AND GARY A. HEWER

Abstract—The sensitivity of the stable discrete-time Lyapunov equation is analyzed through the spectral norm of the inverse Lyapunov operator. This leads to a directly computable easy-to-interpret sensitivity measure, and also provides insight into the connection between sensitivity, stability radius, and conditioning of the eigenproblem of the open-loop state matrix. These results are an extension, to the discrete-time case, of analogous results for the continuous-time Lyapunov equation.

I. Introduction

Properties of the Lyapunov equation are frequently investigated through the associated Kronecker operator. When analyzing the sensitivity of the equation, this approach leads to some measures of conditioning which are difficult to interpret in terms of the system parameters. Moreover, due to the size of the Kronecker product matrices ($n^2 \times n^2$ where n is the order of the matrices in the Lyapunov equation itself), the evaluation of these condition numbers may be problematic for large systems.

Recently, a new sensitivity measure was introduced [10] for the stable continuous-time Lyapunov equation

$$A^H X + XA = -W. \tag{1.1}$$

In (1.1), the unknown X and the matrices A and W are in $C^{n \times n}$ (the space of $n \times n$ complex-valued matrices), A is stable, and A^H denotes the Hermitian transpose of A: $A^H = \bar{A}^T$. No further assumption is made on W. Before getting into more detail, we need the following definitions. A linear mapping Θ over $C^{n \times n}$, defined by

$$\Theta(X) = A^H X + XA \tag{1.2}$$

is called a continuous-time Lyapunov operator. Throughout this paper, the vector norm will be the Euclidean norm, i.e., $\|x\|^2 = x^T x$. Recall the definitions of the Frobenius norm and the spectral norm (also called 2-norm) of a matrix as, respectively,

$$\|M\|_F = \left(\sum_{ij} |m_{ij}|^2\right)^{\frac{1}{2}} = (\operatorname{trace}(M^H M))^{\frac{1}{2}}$$

$$\|M\|_2 = \max_{\|x\|=1} \|Mx\| = \sigma_{\max}(M)$$

where $\sigma_{\max}(M)$ stands for the largest singular value of M. Finally, these two norms induce corresponding norms on the space of linear operators over $C^{n \times n}$, which are, respectively,

$$\|\Theta\|_F = \max_{\|M\|_F = 1} \|\Theta(M)\|_F$$

and

$$\|\Theta\|_2 = \max_{\|M\|_2 = 1} \|\Theta(M)\|_2.$$

In [10], the spectral norm of the inverse Lyapunov operator, given by

$$\|\Theta^{-1}\|_2 = \max_{X \neq 0} \frac{\|X\|_2}{\|A^H X + XA\|_2}$$

is shown to be a relevant measure of the sensitivity of (1.1). Furthermore, this norm turns out to be equal to the spectral norm of the solution H of (1.1), obtained when $W = I$. It also has a simple interpretation in terms of the open-loop system characteristics, namely, in terms of the L_2-norm of the minimally damped solution of $\dot{z} = Az$. More precisely

$$\|H\|_2 = \max_{\|z_0\|=1} \int_0^{+\infty} \|e^{At} z_0\|^2 \, dt.$$

In this paper, we extend these continuous-time results to the discrete case, namely, the discrete-time Lyapunov equation

$$\Omega(X) = Q, \qquad \text{where } \Omega(X) = X - F^H XF. \tag{1.3}$$

The paper is organized as follows. In the next section, we review some basic concepts concerning (1.3) and perform a perturbation analysis in both Frobenius and spectral norm. For each norm, it is shown that the corresponding norm of the operator Ω^{-1} is a relevant measure of the conditioning of (1.3). Some classical techniques of estimating the Frobenius norm of Ω^{-1} are then presented in Section III, before introducing in Section IV our new sensitivity results, based on the evaluation of the spectral norm of Ω^{-1}. Not only does this norm have a closed-form expression in terms of F, but also it can be calculated exactly by solving a single Lyapunov equation. Section V provides some interpretation of $\|\Omega^{-1}\|_2$ in terms of the system natural damping, and gives various bounds relating this norm to characteristics of the matrix F. In light of these results, we discuss in Section VI which characteristics of F are crucial for the conditioning of (1.3). Finally, our conclusions are illustrated by a few selected numerical examples in Section VII.

II. The Lyapunov Equation for Discrete-Time Systems

For continuous-time systems described by the equation $\dot{x} = Ax + Bu$, the stability of the matrix A can be ascertained via (1.1). There is an equivalent result for discrete-time systems $x_{k+1} = Fx_k + Gu_k$. The corresponding Lyapunov stability criterion involves what we call the discrete-time Lyapunov operator, defined as

$$\Omega(X) = X - F^H XF \tag{2.1}$$

where the matrices F and X are in $C^{n \times n}$. In the sequel, we shall call any equation of the type

$$X - F^H XF = Q$$

Manuscript received December 6, 1988; revised November 17, 1989 and April 2, 1990. Paper recommended by Associate Editor, R. V. Patel. This work was supported by the National Science Foundation (and AFOSR) under Grant ECS87-18897, by the National Science Foundation under Grant DMS88-00817, and by the Air Force Office of Scientific Research under Contract AFOSR-89-0167.

P. M. Gahinet is with the National Institute for Research in Computer and Control Sciences (INRIA), Rocquencourt, France.

A. J. Laub and C. S. Kenney are with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106.

G. A. Hewer is with the Naval Weapons Center, China Lake, CA 93555.

IEEE Log Number 9038612.

0018-9286/90/1100-1209$01.00 © 1990 IEEE

[2]

International Journal of High Speed Computing, Vol. 2, No. 2 (1990) 181–191

A PARALLEL ALGORITHM FOR THE MATRIX SIGN FUNCTION

PRADEEP PANDEY, CHARLES KENNEY and ALAN J. LAUB
Department of Electrical and Computer Engineering
University of California
Santa Barbara, CA 93106

Received May 21, 1990

ABSTRACT

We propose a new parallel algorithm for computing the sign function of a matrix. The algorithm is based on the Padé approximation of a certain hypergeometric function which in turn leads to a rational function approximation to the sign function. Parallelism is achieved by developing a partial fraction expansion of the rational function approximation since each fraction can be evaluated on a separate processor in parallel. For the sign function the partial fraction expansion is numerically attractive since the roots and the weights are known analytically and can be computed very accurately. We also present experimental results obtained on a Cray Y-MP.

Keywords: Matrix sign function, parallel algorithms, Padé approximation.

1. Introduction. For a complex scalar z, with $Re(z) \neq 0$, the sign function $sgn(z)$ is defined as:

$$sgn(z) := \begin{cases} +1 & when \;\; Re(z) > 0 \\ -1 & when \;\; Re(z) < 0. \end{cases}$$

This definition can be extended to matrices $X \in \mathbf{C}^{p \times p}$, whose eigenvalues do not lie on the imaginary axis in the following way [14,4]. Let $X = T^{-1}(D + N)T$, where T is nonsingular, $D = \mathrm{diag}(d_1, \ldots, d_p)$, and N is nilpotent and commutes with D. Define the sign of X by

$$sgn(X) = T \,\mathrm{diag}[\, sgn(d_1), \ldots, sgn(d_p)\,]\; T^{-1}.$$

Because the columns of $I - sgn(X)$ and $I + sgn(X)$ form bases of respectively the left-half-plane and right-half-plane invariant subspaces of X, the matrix

TABLE 2
Timing comparison for large test matrices.

size	Newton's method (sec)	Rational method (sec)
98	1.1423	0.2675
298	3.4565	1.3390
398	8.0332	3.1150

can be a problem if the denominator polynomial is ill-conditioned and the weights are large and of differing signs. However, we have also shown that for the sign function, the roots and weight are known analytically and can be computed accurately. Hence, the partial fraction expansion is numerically attractive for the sign function. Experimental results indicate that the parallel algorithm achieves the expected speedup when implemented on a parallel machine like a Cray Y-MP.

Acknowledgement. This research was supported in part by the National Science Foundation (and AFOSR) under Grant No. ECS87-18897, by the National Science Foundation under Grant No. DMS88-00817, by the Air Force Office of Scientific Research under Contract No. AFOSR-89-0167, and by the General Electric Company and the UC MICRO Program (89-096).

REFERENCES

[1] *NAS User Guide*, NAS Systems Division, NASA Ames Research Center, Moffet Field, CA, 1990.

[2] G. A. BAKER, *Essentials of Padé Approximants*, Academic Press, New York, 1975.

[3] G. A. BAKER AND P. GRAVES-MORRIS, *Encyclopedia of Mathematics and its Applications: Padé Approximants, Parts I & II*, Addison-Wesley, Reading, Mass., 1981.

[4] G. J. BIERMAN, *Computational aspects of the matrix sign function solution to algebraic Riccati equation*, in Proc. IEEE Conf. Decis. Control, Las Vegas, Nevada, December 1984, pp. 514–519.

[5] R. BYERS, *Solving the algebraic Riccati equation with the matrix sign function*, Linear Algebra Appl., 85 (1987), pp. 267–279.

[6] E. D. DENMAN AND A. N. BEAVERS, *The matrix sign function and computation in systems*, Appl. Math. Comput., 2 (1976), pp. 63–94.

[7] E. GALLOPOLOUS AND Y. SAAD, *On the parallel solution of parabolic equations*, in Proc. ACM/SIGARCH Int. Conf. on Supercomputing, Crete, Greece, 1989, pp. 17–28.

[8] ——, *Parallel block cyclic reduction algorithm for the fast solution of elliptic equations*, Parallel Comput., 10 (1989), pp. 143–160.

[9] J. D. GARDINER AND A. J. LAUB, *A generalization of the matrix-sign-function solution for algebraic Riccati equations*, Int. J. Control, 44 (1986), pp. 823–832.

[10] P. HENRIC
Sons, I
[11] C. KENNE
to app
[12] A. J. LAI
Auton
[13] ——, *Eff*
Contr
[14] J. D. ROE
by use
[15] G. SZEGÖ
1939.

[3]

SIAM J. CONTROL AND OPTIMIZATION
Vol. 28, No. 6, pp. 1461-1480, November 1990

011

COMPUTABLE BOUNDS FOR THE SENSITIVITY OF THE ALGEBRAIC RICCATI EQUATION*

P. GAHINET† AND A. J. LAUB†

Abstract. In control or estimation theory, linear-quadratic optimization problems give rise to the so-called matrix algebraic Riccati equation (ARE). For such problems, a crucial issue is the existence and uniqueness of a symmetric nonnegative definite stabilizing solution to the ARE, and conditions on the equation parameters are known which guarantee both. However, in the context of computations in finite precision arithmetic, and with imperfect parameter identification, it is of concern whether the ARE retains such a solution in the proximity of a given set of parameters, and how sensitive this solution is to parameter variation.

In this paper, topological properties, such as openness of the domain of existence and continuity with respect to parameters, are established for the symmetric nonnegative definite stabilizing solution. Computable sensitivity estimates are also derived, which quantitatively define a region of safe computation, in terms of the parameters of the equation.

Key words. Riccati equation, sensitivity, stabilizability, computable bounds

AMS(MOS) subject classifications. 49E30, 93B35, 93B40

1. Introduction. The symmetric algebraic Riccati equation (ARE) arises frequently in control and estimation problems. Consider the continuous-time ARE given by:

$$A^TX + XA - XFX + G = 0 \tag{1.1}$$

where all terms are matrices in $\mathbf{R}^{n\times n}$ (real square matrices of order n), and F and G are symmetric, nonnegative definite. The case of complex-valued matrices is qualitatively similar to the sequel but only the real-valued case will be considered here since it is most commonly encountered in applications. Under the assumption that the pairs (A, F) and (G, A) are stabilizable and detectable, respectively, there is a unique nonnegative definite symmetric stabilizing solution X to (1.1) (see [3] or [12]). By X stabilizing (for the pair (A, F)), we mean that $A-FX$ is stable, i.e., all its eigenvalues have strictly negative real parts.

Numerical algorithms are now available that solve the ARE efficiently and dependably, provided the original problem is sufficiently well-conditioned (see [13] or [1]). Well-conditioned means that the solution X is not greatly affected by small perturbations of the data A, F, G. In that case, and with an appropriate scaling of the data (cf. [8]), the Schur-type solvers yield accurate solutions to (1.1).

A natural question following this preliminary remark is how to assess the conditioning of the symmetric ARE, that is, its sensitivity to perturbations of the data. In other words, if we consider a perturbed version of (1.1):

$$(A+\Delta A)^TS + S(A+\Delta A) - S(F+\Delta F)S + G + \Delta G = 0, \tag{1.2}$$

under what conditions does (1.2) keep a unique, nonnegative definite stabilizing solution S? And can we estimate the maximum discrepancy $\|X-S\|$ for a given range of data perturbations ΔA, ΔF, ΔG?

* Received by the editors April 17, 1989; accepted for publication (in revised form) December 13, 1989.

† Department of Electrical and Computer Engineering, University of California, Santa Barbara, California 93106. This research was supported by National Science Foundation grant ECS87-18897 and Air Force Office of Scientific Research contract AFOSR-89-0167.

FA-3 9:00

Proceedings of the 29th Conference
on Decision and Control
Honolulu, Hawaii • December 1990

[4]

ALGORITHMS FOR COMPUTING THE OPTIMAL $\mathcal{H}_\infty$ NORM

Pradeep Pandey * Charles Kenney † Alan J. Laub ‡
Dept. Electrical and Computer Engineering
University of California
Santa Barbara, CA 93106

Andy Packard ‡
Dept. Mechanical Engineering
University of California
Berkeley, CA 94720

Abstract

We present a gradient method for computing the optimal norm for a general $\mathcal{H}_\infty$ control problem. This method is much faster than a bisection method and the additional cost of computing the gradient is small. Convergence is predicated on the smoothness of the spectral radius of the product of certain Riccati solutions. Hybrid bisection-gradient methods can be used in the nonsmooth case.

1 Introduction

Consider the $\mathcal{H}_\infty$ problem described in [1, 2]. We are interested in the following linear system:

$$
\begin{aligned}
\dot{x} &= Ax + B_1 w + B_2 u, \\
z &= C_1 x + D_{11} w + D_{12} u, \\
y &= C_2 x + D_{21} w + D_{22} u.
\end{aligned} \tag{1}
$$

The problem is to find a controller which is internally stabilizing and which minimizes $\gamma := \|T_{zw}\|_\infty$, the closed-loop gain from w to z. We assume that (A, B_1) and (A, B_2) are stabilizable, and that (A, C_1) and (A, C_2) are detectable. These assumptions guarantee that the related $\mathcal{H}_2$ problem is well defined. Further, consider the following assumptions:

A1. $D_{22} = 0$, $D_{12}^T[C_1 \;\; D_{12}] = [\,0 \;\; I\,]$, $D_{21}\,[B_1^T \;\; D_{21}^T] = [\,0 \;\; I\,]$.
A2. $D_{11} = 0$.

These assumptions are from [1] where it was noted that a given system could be transformed so that it meets these assumptions via "loop-shifting" [6]. However, the tranformations to make $D_{11} = 0$ depend on the gain parameter γ. This makes it harder to characterize the behavior of the Riccati solutions as a *function* of γ and hence, is not suitable for the gradient method that we present. Therefore, we will work directly with the more difficult case in which $D_{11} \neq 0$. We will refer to a system that satisfies all the assumptions to be in the *standard* form. A system for which $D_{11} \neq 0$ will be referred to as in the *general* form. For notational simplicity we define $\alpha := \gamma^{-2}$. Scalar- and matrix-valued functions of α will be subcripted with α. If f is such a function then f' and f'' will denote its first and second derivative, respectively, with respect to α.

Let $D_H := \left[\begin{array}{cc} D_{11} & D_{12} \end{array}\right]$ and $M := \left[\begin{array}{cc} D_H^T C_1 & B^T \end{array}\right]^T$, and define the Hamiltonian H_α by

$$
H_\alpha := \left[\begin{array}{cc} A & 0 \\ -C_1^T C_1 & -A^T \end{array}\right] - \mathcal{J} M^T R_\alpha^{-1} M =: \left[\begin{array}{cc} A_\alpha & -F_\alpha \\ -G_\alpha & -A_\alpha^T \end{array}\right], \tag{2}
$$

where

$$
\mathcal{J} := \left[\begin{array}{cc} 0 & I \\ -I & 0 \end{array}\right], \quad R_\alpha := D_H^T D_H - \left[\begin{array}{cc} \alpha^{-1} I & 0 \\ 0 & 0 \end{array}\right]. \tag{3}
$$

The Hamiltonian J_α is defined similarly. The following theorem is essentially from [1].

Theorem 1 *There exists an admissible controller such that $\|T_{zw}\|_\infty < \gamma$ iff the following three conditions hold:*

(i) $H_\alpha \in dom(Ric)$ and $X_\alpha := Ric(H_\alpha) \geq 0$,
(ii) $J_\alpha \in dom(Ric)$ and $Y_\alpha := Ric(J_\alpha) \geq 0$,
(iii) $\rho_\alpha := \rho(X_\alpha Y_\alpha) < \alpha^{-1}$.

In the sequel we describe the behavior of H_α and X_α. By duality, these comments also apply to J_α and Y_α. Consider the Riccati equation associated with H_α in (2)

$$
\mathcal{R}(X, \alpha) := A_\alpha^T X_\alpha + X_\alpha A_\alpha - X_\alpha F_\alpha X_\alpha + G_\alpha = 0. \tag{4}
$$

By an *admissible* solution we will mean a unique positive semi-definite stabilizing solution. For $\alpha = 0$ our assumptions guarantee *a priori* that an admissible $X_0 := Ric(H_0) \geq 0$ exists. However, for $\alpha > 0$ it is not clear for what values of α an admissible solution will exist. By continuity, H_α will be in *dom(Ric)* for some $\alpha > 0$. Suppose H_α and J_α first fail to be in *dom(Ric)* at α_x and α_y, respectively, and define $\alpha_* := \min(\alpha_x, \alpha_y)$. We will refer to an $\alpha \in (0, \alpha_*)$ as *feasible*. The following theorems, which are proved in [5], characterize the behavior of H_α and X_α for a general problem.

Theorem 2 *For feasible α we can write $G_\alpha = C_\alpha^T C_\alpha$. If $\mathcal{V}$ spans the unobservable subspace of the pair (A, C_1) then it also spans the unobservable subspace (A_α, C_α). Further, $\ker(X_\alpha) = \mathcal{V}$.*

Theorem 3 *Let $\alpha_* > \alpha_1 > \alpha_0 \geq 0$. Then for a general problem the Riccati solutions X_α and Y_α, and the spectral radius function ρ_α are continuous and nondecreasing, i.e., $X_{\alpha_1} \geq X_{\alpha_0}$, and $\rho_{\alpha_1} \geq \rho_{\alpha_0}$.*

2 Gradient Method

In this section we review a gradient method for computing γ_{opt} [4]. We will assume that condition (*iii*) of Theorem 1 will fail before condition (*i*) or (*ii*). This implies that at the optimal value $\rho(X_\alpha Y_\alpha) = \frac{1}{\alpha}$, and that α_{opt} is a root of the equation

$$
h(\alpha) := \alpha\rho_\alpha - 1 = 0. \tag{5}
$$

We can find the root using the Halley-secant (or Newton's) method. The derivatives of $h(\alpha)$ are

$$
h' = \alpha\rho', \quad \text{and} \quad h'' = 2\rho' + \rho'', \tag{6}
$$

assuming that the derivatives of the spectral radius function exist. Using these definitions the Halley-secant method yields

$$
\alpha_1 = \alpha_0 + \Delta\alpha, \quad \text{where} \quad \Delta\alpha = -h_{\alpha_0}\left[h' - \frac{h'' h_{\alpha_0}}{2h'}\right]^{-1}. \tag{7}
$$

*Supported in part by National Science Foundation (and AFOSR) under Grant No. ECS87-18897.

†Supported in part by the National Science Foundation under Grant No. DMS88-00817 and the Air Force Office of Scientific Research under Contract No. AFOSR-89-0167.

‡Supported in part by the General Electric Company and the UC MICRO Program (89-096).

[5]

Numerical Solution of Large-Scale Riccati Equations

PRADEEP PANDEY, CHARLES KENNEY, and ALAN J. LAUB
Department of Electrical and Computer Engineering
University of California
Santa Barbara, CA 93106

Abstract

We discuss algorithms for obtaining numerical solutions for large-scale algebraic Riccati equations. By "large-scale" we mean problems arising from models involving matrices of dimensions in the hundreds or perhaps thousands. There are numerous sources for such problems, e.g., control of large space structures, distributed parameter systems, and interconnected power systems. Design of control systems for such large systems places a significant burden on computing resources and may require unacceptably long computing time on existing computers. We describe a parallel algorithm for solving Riccati equations which is based on the matrix sign function. We also present numerical results obtained on a Cray Y-MP for a structures model currently under study by Rockwell.

1 Introduction

Algebraic Riccati equations (ARE) play a central role in control theory. Indeed, they are one of the most deeply studied nonlinear matrix equations arising in mathematics and engineering. Riccati equations arise in a variety of situations and their role in control and system theory is well established.

Consider the following state-space model of a linear system:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t), \end{aligned} \tag{1}$$

where the state $x(t) \in \mathbb{R}^n$, the input $u(t) \in \mathbb{R}^m$, and the output $y(t) \in \mathbb{R}^p$. In an LQG problem we define an associated quadratic cost function

$$J(u) := \int_0^{+\infty} (y^T y + u^T u)\, dt,$$

and we want to find a state-feedback control u such that the closed-loop system is stable and the cost function J is minimized. This optimization problem leads to the following ARE:

$$A^T X + XA - XBB^T X + C^T C = 0, \tag{2}$$

where $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, and $C \in \mathbb{R}^{p\times n}$. Natural assumptions of stabilizability and detectability guarantee existence of $X = X^T \in \mathbb{R}^{n\times n}$ which is the unique nonnegative definite stabilizing solution of (2).

We are interested in efficient algorithms for solving large-scale AREs. Our primary motivation is that in the future control engineers can increasingly be expected to be interested in solving problems of large dimensions. At present, problem sizes of "tens" or even low "hundreds" can be handled by

Size of problem	542
Time (seconds)	207
One-norm of residual $\mathcal{R}$	3.9093×10^{-9}

Unfortunately, we cannot deduce accuracy of the computed solution from the norm of the residual alone. Suppose S is the true admissible solution for the ARE in (2). To get a better measure of accuracy we use the following bound from Kenney *et al.* [14]:

$$\|X - S\| \leq 2\|\Omega^{-1}\|\|\mathcal{R}\|, \tag{37}$$

where Ω is the closed-loop Lyapunov operator defined by

$$\Omega(Z) := A_c^T Z + Z A_c, \quad \text{where} \quad A_c := A - BB^T X.$$

In [14] Kenney *et al.* also showed that $\|\Omega^{-1}\| = \|Z\|$ where $Z \in \mathbb{R}^{r \times r}$ solves

$$A_c^T Z + Z A_c + I = 0. \tag{38}$$

We solved the above equation to obtain $\|Z\| = 1.362 \times 10^3$. From the data we conclude that the solution X was computed to at least 6-digit accuracy.

6 Summary

We have proposed a new parallel algorithm for solving large-scale Riccati equations. The algorithm is based on a partial fraction expansion of a certain Padé approximation used in computing the matrix sign function. A partial fraction expansion allows a parallel implementation in which each fraction can be evaluated on a different processor. In general, the roots and weights of such an expansion have to be computed numerically. This can be a problem if the denominator polynomial is ill-conditioned and the weights are large and of differing signs. However, we have also shown that for the sign function, the roots and weights are known analytically and can be computed accurately. Hence, the partial fraction expansion is numerically attractive for the sign function. Experimental results indicate that the parallel algorithm achieves the expected speedup when implemented on a parallel machine like a Cray Y-MP.

Acknowledgements

This research was supported in part by the National Science Foundation (and AFOSR) under Grant No. ECS87-18897, by the Air Force Office of Scientific Research under Contract No. AFOSR-89-0167, and by Rockwell International's University Control Technology Program.

References

[1] Baker, G.A., *Essentials of Padé Approximants*, Academic Press, New York, 1975.

[2] Baker, G.A. and P. Graves-Morris, *Encyclopedia of Mathematics and its Applications: Padé Approximants, Parts I & II*, Addison-Wesley, Reading, Mass., 1981.

[3] Barraud, A.Y., "Investigation autour de la fonction signe d'une matrice—application à l'équation de Riccati," *R.A.I.R.O. Automatique/Systems Analysis and Control*, 13 (1979), pp. 335–368.

[6]

SIAM J. MATRIX ANAL. APPL.
Vol. 12, No. 2, pp. 273–291, April 1991

006

RATIONAL ITERATIVE METHODS FOR THE MATRIX SIGN FUNCTION*

CHARLES KENNEY† AND ALAN J. LAUB†

Abstract. In this paper an analysis of rational iterations for the matrix sign function is presented. This analysis is based on Padé approximations of a certain hypergeometric function and it is shown that local convergence results for "multiplication-rich" polynomial iterations also apply to these rational methods. Multiplication-rich methods are of particular interest for many parallel and vector computing environments. The main diagonal Padé recursions, which include Newton's and Halley's methods as special cases, are globally convergent and can be implemented in a multiplication-rich fashion which is computationally competitive with the polynomial recursions (which are not globally convergent). Other rational iteration schemes are also discussed, including Laurent approximations, Cayley power methods, and globally convergent eigenvalue assignment methods.

Key words. Padé approximation, matrix sign function, Riccati equations, rational iterations

AMS(MOS) subject classifications. 15A24, 65D99, 65F99

1. Introduction. It is a classical result that the algebraic Riccati equation can be solved by using an invariant subspace of an associated Hamiltonian matrix. This motivated the introduction, by Roberts [21] in 1971, of the matrix sign function as a means of finding the positive and negative invariant subspaces of any matrix X which does not have eigenvalues on the imaginary axis. This and subsequent work [9] showed that the matrix sign function could be used to solve many problems in control theory.

The sign of X can be defined constructively as the limit of the Newton sequence

$$X_{n+1} = \tfrac{1}{2}(X_n + X_n^{-1}), \qquad X_0 = X, \tag{1.1}$$

$$\operatorname{sgn}(X) = \lim_{n \to +\infty} X_n. \tag{1.2}$$

Newton's method has the pleasant feature that it is globally convergent; if X has no eigenvalues on the imaginary axis then the limit in (1.2) exists. As a definition, however, (1.2) does not reveal many of the important properties of the sign function. Because of this, it is useful to have an equivalent definition based on the Jordan canonical form of X (see [4], [7]). For a complex scalar z with $\operatorname{Re} z \neq 0$, define the sign of z by

$$\operatorname{sgn} z = \begin{cases} 1 & \text{if } \operatorname{Re} z > 0, \\ -1 & \text{if } \operatorname{Re} z < 0. \end{cases} \tag{1.3}$$

For a complex matrix X such that $\Lambda(X) \subset \mathbf{C}^+ \cup \mathbf{C}^-$ (i.e., X has no eigenvalues on the imaginary axis) let T take X to Jordan form:

$$X = T^{-1}\begin{bmatrix} P & 0 \\ 0 & N \end{bmatrix} T, \tag{1.4}$$

* Received by the editors September 28, 1989; accepted for publication (in revised form) November 15, 1989. This research was supported by National Science Foundation (and Air Force Office of Scientific Research) grant ECS87-18897, National Science Foundation grant DMS88-00817, and Air Force Office of Scientific Research contract AFOSR-89-0167.

† Department of Electrical and Computer Engineering, University of California, Santa Barbara, California 93106 (laub%lanczos@hub.ucsb.edu).

[7]

SIAM J. SCI. STAT. COMPUT.
Vol. 12, No. 3, pp. 488-504, May 1991

POLAR DECOMPOSITION AND MATRIX SIGN FUNCTION CONDITION ESTIMATES*

CHARLES KENNEY† AND ALAN J. LAUB†

Abstract. This paper presents reliable condition estimation procedures, based on Fréchet derivatives, for polar decomposition and the matrix sign function. For polar decomposition, the condition number for complex matrices is equal to the reciprocal of the smallest singular value, and rather surprisingly, for real matrices it is equal to the reciprocal of the average of the two smallest singular values. By using inverse power methods, both of these condition numbers can be evaluated at a fraction of the cost of finding the polar decomposition.

Except for special cases, such as for normal matrices, the condition number of the matrix sign function does not have such a precise characterization. However, accurate condition estimates can be obtained by using explicit forms of the Fréchet derivative, or its finite-difference approximation, with a matricial inverse power method. These methods typically require two extra sign function evaluations, and it is an open problem whether accurate estimates can be obtained for a fraction of a function evaluation, as is the case for the polar decomposition. Related results for the stable Lyapunov equation and Newton's method for the matrix square root problem are discussed.

Key words. polar decomposition, matrix sign function, conditioning

AMS(MOS) subject classifications. 65F35, 15A12

1. Introduction. Both the polar decomposition and the matrix sign function play important roles in many matrix algorithms [1]-[3], [8], [9], [14], [17]. The sensitivity of these matrix functions is determined by the norms of their Fréchet derivatives. More specifically, let $F = F(X)$ be a matrix function which is continuously differentiable at X in the sense that there exists a linear matrix operator $L(\cdot) = L(\cdot, X, F)$ such that for any matrix Z,

$$\lim_{\delta \to 0} \frac{F(X+\delta Z) - F(X)}{\delta} = L(Z). \tag{1.1}$$

Then L is the Fréchet derivative of F at X and we define the absolute and relative condition numbers of F at X, with respect to a matrix norm $\|\cdot\|$, by $K_a(F,X) = \|L\| = \max_{\|Z\|\neq 0} \|L(Z)\|/\|Z\|$, and $K_r(F,X) = \|L\|\,\|X\|/\|F(X)\|$ when $\|F(X)\| \neq 0$. This definition is consistent with the condition theory of Rice [16] (see [13]). For δ small and $\|Z\| = 1$, we see from (1.1) that

$$\|\Delta F\| = \|F(X+\delta Z) - F(X)\| \approx \|L(Z)\|\delta \leq \|L\|\delta,$$

which motivates the use of $\|L\|$ as a condition measure for the mapping $X \to F(X)$.

As a simple example, if $F(X) = X^2$ then

$$\frac{F(X+\delta Z) - F(X)}{\delta} = ZX + XZ + \delta Z^2,$$

so $L(Z) = ZX + XZ$ and $\|L\| \leq 2\|X\|$.

* Received by the editors June 21, 1989; accepted for publication (in revised form) December 4, 1989. This research was supported by National Science Foundation (and Air Force Office of Scientific Research) grant ECS 87-18897, National Science Foundation grant DMS88-00817, and Air Force Office of Scientific Research contract AFOSR-89-0167.

† Department of Electrical and Computer Engineering, University of California, Santa Barbara, California 93106.

[8]

Square Root Parallel Kalman Filtering Using Reduced-Order Local Filters

S. ROY
R. H. HASHEMI
A. J. LAUB, Fellow, IEEE
University of California at Santa Barbara

We examine various discrete-time parallel Kalman filtering implementations, with special attention given to square root versions in both covariance and information filter forms. Throughout the paper we employ the common convention of using the term "square root" to refer to a Cholesky factor. A special feature of the suggested architecture is the ability to accommodate parallel local filters that have a smaller state dimension than the global filter. The estimates and covariance or information matrices (or their square roots) from these reduced-order filters are collated at a central filter at each step to generate the full-size, globally optimal estimates and their associated error covariance or information matrices (or their square roots).

Manuscript received February 26, 1990; revised August 28, 1990.

IEEE Log No. 41908.

This work was supported by the Air Force Office of Scientific Research under Contract AFOSR-89-0167 and by the National Science Foundation (and AFOSR) under Grant ECS87-18897.

Authors' current addresses: S. Roy, Dept. of Electrical Engineering, University of Pennsylvania, 200 S. 33rd St., Philadelphia, PA 19104; R. H. Hashemi, Dept. of Radiology, Kaiser Permanente Center, Los Angeles, CA 90027; A. J. Laub, Dept. of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106.

I. INTRODUCTION

Decentralized estimation problems have been the focus of great interest in recent years in multisensor environments such as integrated navigation systems. This stems from the fact that using a monolithic centralized filter in a multisensor system can result in severe computational problems due to overloading the filter with more data than it can handle. Consequently, the overall centralized system may be unreliable and suffer from poor accuracy and stability. This leads naturally to decentralized processing configurations.

Various decentralized and parallel versions of the Kalman filter have been reported in the literature [1–10]. Conventional Kalman filtering algorithms, though well known to be theoretically globally optimal (e.g., [11]), can also be numerically unreliable. To remedy such problems, more numerically stable and better conditioned implementations of the Kalman filtering algorithms, such as U-D or square root formulations, are employed [12, 13]. Recent efforts have concentrated on decentralized versions of such numerically stable filtering algorithms, e.g., [14–18].

Decentralized estimation offers numerous advantages in many applications. Indeed, in many circumstances it provides the most logically feasible processing schemes. For instance, in a multisensor system in which each individual sensor has its own "built-in" Kalman filter, one is interested in combining the estimates from these independent data sources (i.e., the built-in Kalman filters) to generate a global estimate that will, ideally, be optimal. Furthermore, decentralization makes for easy fault detection and isolation [5], since the output of each local sensor filter can be tested and, if a sensor should fail, it can be expeditiously removed from the sensor network before it affects the aggregate filter output. Also, decentralization increases the input data rates significantly and yields moderate improvements in the throughput.

The focus of several existing parallel Kalman filter structures ([1–3, 6, 7, 9, 10]) has been to preserve the overall global optimality of the whole system, which is definitely a desirable feature and serves as a benchmark for other systems. While all the previous results have assumed *full-order* local models to achieve global optimality, the intent of this work is to investigate scenarios and present algorithms for which the same can be achieved using *reduced-order* local filters. In general, this cannot be achieved for arbitrary global and (reduced-order) local models; more is said about this in Section II. Although this limits the applicability of the decentralized schemes proposed here, there are nevertheless several useful scenarios for which these algorithms are natural candidates. These include all cases when the global state vector can be partitioned into *disjoint* segments, and each segment or subvector yields a compatible reduced-order local

[9]

7 Invariant Subspace Methods for the Numerical Solution of Riccati Equations

Alan J. Laub

7.1 Introduction

In this tutorial paper, an overview is given of progress over the past ten to fifteen years towards reliable and efficient numerical solution of various types of Riccati equations. Our attention will be directed primarily to matrix-valued algebraic Riccati equations and numerical methods for their solution based on computing bases for invariant subspaces of certain associated matrices. Riccati equations arise in modeling both continuous-time and discrete-time systems in a wide variety of applications in science and engineering. One can study both algebraic equations and differential or difference equations. Both algebraic and differential or difference equations can be further classified according to whether their coefficient matrices give rise to so-called symmetric or nonsymmetric equations. Symmetric Riccati equations can be further classified according to whether or not they are definite or indefinite.

The rest of the paper is organized as follows. A brief review of "classical" methods is followed by a summary of the now-standard Schur method, introduced in 1978, for solving algebraic Riccati equations. Extensions of the basic Schur method, by means of associated generalized eigenvalue problems, are then described together with some applications. Next, some powerful new numerical results relating to Riccati equations in general will be described. These include a thorough analysis of iterative refinement via Newton's method (including a computable estimate of the region of convergence), a theorem on the relation of error estimates to residuals, estimation of the condition of algebraic Riccati equations, and promising new scaling strategies. Newton's method for computing the matrix sign function is then described and its implementation for parallel algorithms for Riccati equations (on a message-passing hypercube computer) is outlined. This method is particularly well suited to parallelization and vectorization and has been used successfully to solve fairly large order (several hundred) problems. A number of generalizations of this basic iteration have extended its applicability to a broader range of problems. For example, generalizations of the matrix sign function to the case of matrix pencils allows straightforward solution of discrete-time Riccati equations. Furthermore, the Newton iteration itself has been generalized considerably and found to be but a special case of a general family of iterations for the matrix

Bittanti, Laub, Willems (Eds.)
The Riccati Equation
© Springer-Verlag 1991

ition of Riccati equa-
iations and invariant
ed on direct methods
natrix sign function.
g Riccati equations.
ten to fifteen years.
.RE) was something
Existing techniques
ethod suggested for
ctorization problem.
polynomial matrices
erally conceded that
state space methods
olving an ARE has

spaces have proven
many hundred. A
mance in the finite
er, many important
research.
related to Riccati

ms may be partic-
s, especially those
certainly in 1978
between the alge-
ces leads naturally
larities which pre-
le in, for example
s, and Mehrmann,

symmetric Riccati
s nowhere near as
etric case, we are
which arise in the
e game-theoretic

increasing study
er modeling with
me and transient
for stiff Riccati
continuing study.

- software: This is almost always, of course, the ultimate vehicle of reliable technology transfer. Early attempts at a large comprehensive Fortran-based Riccati package (RICPACK; see [11]) will undoubtedly be superseded by much more easily constructed packages based on software such as MATLAB and its clones and imitators.

Acknowledgements. Much of the research surveyed in this paper represents collaborative efforts with past and present students including Thrasyvoulos Pappas, Kwae Hi Lee, Bo Arnold, Doug Bender, Judy Gardiner, Matt Wette, Chiu Choi, Pascal Gahinet, Pradeep Pandey, Thorkell Gudmundsson, and Ali Ghavimi. It has also been a special privilege to have the opportunity to collaborate with my colleague, Charles Kenney, of the Scientific Computation Laboratory in the Electrical and Computer Engineering Department at UCSB.

We are grateful for research support from the National Science Foundation (and AFOSR) under Grant No. ECS87-18897, the National Science Foundation under Grant No. DMS88-00817, and the Air Force Office of Scientific Research under Contract No. AFOSR-89-0167.

References

1. C. Ahlbrandt. Continued fraction representations of maximal and minimal solutions of a discrete matrix Riccati equation. preprint. (1990).
2. F.A. Aliev, B.A. Bordyug, and V.B. Larin. The spectral method of solving matrix algebraic Riccati equations. *Soviet Math. Dokl.*, 35:121–125, 1987.
3. F.A. Aliev, B.A. Bordyug, and V.B. Larin. Orthogonal projections and the solution of a matrix algebraic Riccati equation. *Soviet Math. Dokl.*, 38:531–534, 1989.
4. B.D.O. Anderson. Solution of quadratic matrix equations. *Electron. Letters*, 2:371–372, 1966.
5. B.D.O. Anderson. Second-order convergent algorithms for the steady-state Riccati equation. *Int. J. Control*, 28:295–306, 1978.
6. B.D.O. Anderson and J.B. Moore. *Linear Optimal Control*. Prentice Hall, Englewood Cliffs, NJ, 1971.
7. B.D.O. Anderson and J.B. Moore. *Optimal Filtering*. Prentice Hall, Englewood Cliffs, NJ, 1979.
8. L.R. Anderson, D.W. Brewer, and A.R. Baykam. Numerical solution of the symmetric Riccati equation through iteration. In *Proc. 1982 Amer. Control Conf.*, pages 1010–1015, Arlington, VA, June 1982.
9. S. Arimoto. Optimal feedback control minimizing the effects of noise disturbances. *Trans. SICE*, 2(1966), pp. 1–7, (in Japanese).
10. W.F. Arnold, III. *On the Numerical Solution of Algebraic Matrix Riccati Equations*. PhD thesis, Univ. of Southern California, Dept. of Elec. Eng.—Systems, Los Angeles, CA, December 1983.
11. W.F. Arnold, III and A.J. Laub. A software package for the solution of generalized algebraic Riccati equations. In *Proc. 22nd IEEE Conf. Decis. Control*, pages 415–417, San Antonio, TX, December 1983.
12. W.F. Arnold, III and A.J. Laub. Generalized eigenproblem algorithms and software for algebraic Riccati equations. *Proc. of IEEE*, 72:1746–1754, 1984.
13. M. Athans, Editor. Special issue on the linear-quadratic-Gaussian estimation and control problem. *IEEE Trans. Auto. Control*, AC-16, December 1971.

[10]

A Gradient Method for Computing the Optimal H_∞ Norm

P. Pandey, C. Kenney, A. Packard, and A. J. Laub

Abstract—One of the key computational issues in the design of H_∞ optimal controllers is the determination of the optimal H_∞ norm. State-space methods to compute this norm depend on solving certain

Manuscript received November 17, 1989; revised June 15, 1990. The work of P. Pandey was supported in part by the National Science Foundation and the Air Force Office of Scientific Research under Grant ECS 87-18897; the work of C. Kenney was supported in part by the National Science Foundation under Grant DMS88-00817 and by the Air Force Office of Scientific Research under Contract AFOSR-89-0167; the work of A. J. Laub and A. Packard was supported in part by the General Electric Company and by the UC MICRO Program 89-096.

The authors are with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106.

IEEE Log Number 9143247.

0018-9286/91/0700-0887$01.00 ©1991 IEEE

[11]

Local Convergence Analysis of Conjugate Gradient Methods for Solving Algebraic Riccati Equations *

Ali R. Ghavimi, Charles Kenney, and Alan J. Laub
Dept. of Electrical and Computer Engineering
University of California
Santa Barbara, CA 93106

Abstract

Necessary and sufficient conditions are given for local convergence of the conjugate gradient (cg) method for solving symmetric and nonsymmetric algebraic Riccati equations. For these problems, the Frobenius norm of the residual matrix is minimized via the cg method, and convergence in a neighborhood of the solution is predicated on the positive definiteness of the associated Hessian matrix. For the nonsymmetric case, the Hessian eigenvalues are determined by the squares of the singular values of the closed-loop Sylvester operator. In the symmetric case, the Hessian eigenvalues are closely related to the squares of the closed-loop Lyapunov singular values. In particular, the Hessian is positive definite if and only if the associated operator is nonsingular. The invertibility of these operators can be expressed as a non-cancellation condition on the eigenvalues of the closed-loop matrices. For example, the stability of the closed-loop matrix, for the positive semi-definite Riccati solution, ensures the invertibility of the Lyapunov operator and hence the convergence of the cg method in a neighborhood of that solution.

1 Introduction

When minimizing a scalar function f via the conjugate gradient (cg) method, local convergence is equivalent to the Hessian of f being positive definite at the point of minimization [1].

*This research was supported in part by the National Science Foundation (and AFOSR) under Grant No. ECS87-18897, by the National Science Foundation under Grant No. DMS88-00817, and by the Air Force Office of Scientific Research under Contract No. AFOSR-89-0167.

[12]

Algebraic Riccati Equations and the Distance to the Nearest Uncontrollable Pair *

P. Gahinet and A.J. Laub

Department of Electrical and Computer Engineering
University of California, Santa Barbara, CA 93106

Abstract

A connection is established between nearness to unstabilizability of a stabilizable pair (A, B) of matrices and nearness to singularity of the symmetric positive definite solution to an associated algebraic Riccati equation. From this result, computable upper and lower bounds are derived for the distance of (A, B) to the nearest uncontrollable pair. Numerical tests confirm the validity of the method and potential applications are discussed.

1 Introduction

When numerically assessing whether a pair of matrices $(A, B) \in \mathbf{R}^{n\times n} \times \mathbf{R}^{n\times r}$ is controllable (or stabilizable), tests which simply provide a yes/no answer are not entirely satisfactory [17, 18]. Instead, an estimate of how far the pair is from the set of uncontrollable (respectively, unstabilizable) pairs is more relevant. Unfortunately, this involves a nonconvex minimization in a space of n dimensions and existing numerical methods to search for minima often suffer from the following limitations:

- the computed minima are only local,
- a two-dimensional search is necessary when complex perturbations are allowed,
- the speed of convergence is guaranteed to be quadratic only in the proximity of the local minima and a high computational overhead may thus be attached.

Few lower or upper bounds on the distance to uncontrollability are available in the literature. Upper bounds were proposed in [1] but they require either forming the controllability matrix, or that A be stable. A lower bound was obtained by Demmel in [6].

*This research was supported by the National Science Foundation (and AFOSR) under Grant No. ECS87-18897 and the Air Force Office of Scientific Research under Contract No. AFOSR-89-0167. Replies should be addressed to the first author, currently with INRIA, Domaine de Voluceau, BP 105, 78153 Le Chesnay Cedex, France.

[13]

On Scaling Newton's Method for Polar Decomposition and the Matrix Sign Function *

Charles Kenney Alan J. Laub
Department of Electrical and Computer Engineering
University of California
Santa Barbara, CA 93106

Abstract

A tight bound is given on the speed of convergence of Newton's method with optimal scaling for the polar decomposition of a nonsingular complex matrix. Necessary and sufficient conditions are then derived which tell when an approximation to the optimal scaling value will give better results than the unscaled Newton method. For the related matrix sign problem, it is shown that optimal scaling requires complete knowledge of the eigenvalues of the original matrix. Because this is impractical, we derive a family of scaling methods which are optimal with respect to partial eigenvalue information. This family includes optimal scaling as well as a 'semi-optimal' scaling method based on the dominant eigenvalues of the matrix and its inverse. Semi-optimal scaling can be implemented using the power method and gave nearly optimal performance on a set of test problems. These test problems also show that a variety of other commonly used scaling strategies, including spectral scaling, determinantal scaling, and 2-norm scaling can result in unduly slow convergence.

Keywords: polar decomposition, matrix sign function, Newton's method, optimal scaling.
AMS(MOS) subject classification: 65F35,65F30,15A18.
Abbreviated title: Scaling Newton's method.

1 Introduction

The polar decomposition of a nonsingular complex matrix A of order m is a matrix pair (U, H) such that U is unitary, H is Hermitian and positive definite, and $A = UH$. If A has a singular value decomposition, $A = P\Sigma Q^H$, where P and Q are unitary and $\Sigma = diag(\sigma_1, \ldots, \sigma_m)$ with $0 < \sigma_m \leq \cdots \leq \sigma_1$, then [7]

$$U = PQ^H,\ H = Q\Sigma Q^H. \tag{1}$$

However, it is more efficient to compute the polar decomposition using scaled Newton recursions of the form,

$$A_{n+1} = \frac{1}{2}(\gamma_n A_n + (\gamma_n A_n^H)^{-1});\ A_0 = A,\ \gamma_n > 0. \tag{2}$$

For γ_n suitably chosen [7], [10], $A_n \to U$, and H can be found from $H = U^H A$. In the next section we give convergence bounds supporting the empirical observation that Newton's method with optimal scaling,

*This research was supported by the National Science Foundation (and AFOSR) under Grant No. ECS87-18897, the National Science Foundation under Grant No. DMS88-00817, and the Air Force Office of Scientific Research under Contract No. AFOSR-89-0167.

[14]

Scaling of the Discrete-Time Algebraic Riccati Equation to Enhance Stability of the Schur Solution Method *

Thorkell Gudmundsson, Charles Kenney, and Alan J. Laub
Dept. of Electrical and Computer Engineering
University of California
Santa Barbara, CA 93106

Abstract

A simple scaling procedure for discrete-time Riccati equations is introduced. This procedure eliminates instabilities which can be associated with the linear equation solution step of the generalized Schur method without changing the condition of the underlying problem. A computable bound for the relative error of the solution of the Riccati equation is also derived.

1. Introduction

The Schur method [8] for solving discrete-time algebraic Riccati equations consists of transforming an associated generalized eigenvalue problem to real Schur form using orthogonal equivalence transformations, followed by the solution of a system of linear equations. The orthogonal transformations are numerically well-conditioned, but recent work [12] has suggested that the overall method can appear numerically unstable, even when the original equation is well-conditioned. This can originate for two different reasons. One is the ill-conditioning of a linear system of equations, and the other is related to scaling problems for the basis vectors of a certain subspace. In this paper we extend the work which was done for the continuous-time Riccati equation in [7] to the discrete-time equation and show that this apparent numerical instability can be eliminated by a scalar scaling procedure. Moreover, this analysis yields a good computable bound on the relative error of the solution of the Riccati problem.

Our procedure is not completely satisfactory, because the scalar involved is a function of the solution to be computed and thus leaves open the question of how to estimate it accurately beforehand. However, this does not invalidate our main result in any way, namely that the Schur method is not inherently numerically unstable. In fact, the problem can be circumvented by solving the unscaled equation, using our error bound to establish the accuracy of the computed solution, and in case that is not satisfactory, using the inaccurate solution to estimate the optimal scaling parameter. This estimate will almost always be sufficiently close to the correct value to yield an accurate result on the second pass.

*This research was supported in part by the National Science Foundation (and AFOSR) under Grant No. ECS87-18897 and by the Air Force Office of Scientific Research under Contract No. AFOSR-89-0167.

[15]

Orthogonal Canonical Forms for Second-Order Systems *

Trevor Williams
Dept. of Aerospace Engineering and Engineering Mechanics
University of Cincinnati
Cincinnati, OH 45221-0070

and

Alan J. Laub
Dept. of Electrical and Computer Engineering
University of California
Santa Barbara, CA 93106

Abstract

It is shown in this paper that a linear second-order system with arbitrary damping cannot be reduced to Hessenberg-triangular form by means of orthogonal transformations. However, it is also shown that such an orthogonal reduction is always possible for the modal damping commonly assumed for models of flexible structures. In fact, it is shown that modally damped models can be orthogonally reduced to a new triangular second-order Schur form.

1 Introduction

Second-order models arise naturally in the study of many types of physical systems, with common examples being electrical and mechanical networks. An application area of great practical interest for dynamics and control is that of flexible space structures (FSS) [2], which are commonly represented by second-order finite element models of very high dimension. Now, continuum models of structures are, to be sure, much more elegant (see, for example, [1, 11]) but it is generally still the case that setting up the governing partial differential equations and solving the resulting boundary value problems can only be done for relatively simple structures. In analyzing a realistic structure (spacecraft, airplane, etc.), a continuous structure model is seldom feasible and common engineering practice has been to use some method (usually finite elements) to get an approximate "M and

*This work was initiated while the first author held a National Research Council-NASA Langley Research Center Senior Research Associateship. The work of the second author was supported by the National Science Foundation (and AFOSR) under Grant No. ECS87-18897 and the Air Force Office of Scientific Research under Contract No. AFOSR-89-0167.